УДК 004.89:004.93

А.А .Николенко, Тьен Т.К. Нгуен
Одесский национальный политехнический университет, МОН МС Украины
Украина, 65044, г. Одесса, просп. Шевченко, 1

Обнаружение текстовых областей в видеопоследовательностях

A.A. Nikolenko, Tien T.K. Nguyen
Odessa National Polytechnic University Ministry of Education, Youth and Sports of Ukraine, c. Odessa
Ukraine, 65044, c. Odessa, Shevchenko av., 1

Text Region Detection in Video Framess

А.О. Ніколенко, Тьєн Т.К. Нгуєн
Одеський національний політехнічний університет МОНмолодьспорту України
Україна, 65044, м. Одеса, просп. Шевченка, 1

Виявлення текстових областей у відеопослідовностях

В статье рассматривается задача обнаружения текстовых областей на неоднородном фоне в видеопоследовательностях. Предлагается двухэтапная схема, алгоритм и методика обнаружения текстовых областей с использованием непрерывного вейвлет-преобразования с автоматическим выбором масштаба или итерационной обработкой с разными масштабами.
Ключевые слова: обнаружение текстовых областей, обработка изображений, анализ изображений.

In the article, the problem of text region detection on the non-uniform background in video frames is considered. The two-stage scheme, the algorithm and the technique of detection of text regions using a continuous wavelet transform with automatic scale selection or iteration processing at different scales are poposed.
Key Words: text region detection, image processing, image analysis.

У статті розглядається задача виявлення текстових областей на неоднорідному фоні в відеопослідовності. Пропонується двоетапна схема, алгоритм і методика виявлення текстових областей з використанням безперервного вейвлет-перетворення з автоматичним вибором масштабу або ітераційною обробкою з різними масштабами.
Ключові слова: виявлення текстових областей, обробка зображень, аналіз зображень.

Введение

Для обработки постоянно возрастающих объемов информации с мультимедийным контентом весьма актуальной является разработка методов и алгоритмов для информационных систем индексации и семантического поиска информации. При обработке видеопоследовательностей часто приходится обнаруживать и распознавать текстовую информацию (например, номера железнодорожных вагонов, автомобилей, контейнеров, надписи на дорожных знаках или других объектах), которая может служить источником семантических признаков различных уровней.

При анализе видеопоследовательностей различают искусственно добавленный текст и текст сцены. Добавленный текст содержит дополнительную информацию, связанную с видеопоследовательностью (например, бегущая строка при просмотре телевизионных новостей). Текст сцены – текстовая область, которая была захвачена камерой как часть сцены (например, надписи на футболках спортсменов или дорожных знаках).

При извлечении текстовой информации из видеопоследовательностей одной из важных процедур является обнаружение текстовых областей. Существует множество алгоритмов и методов для ее выполнения, однако до настоящего времени нет окончательного решения этой задачи.

Большинство из предложенных методов обнаружения текстовых областей используют в качестве признаков текста информацию о цвете, контурах и текстурных особенностях соответствующего изображения. Чтобы воспользоваться этой информацией, то есть описать текст и отличить его от фона, некоторые исследователи применяют эвристические правила, в то время как другие используют метод машинного обучения на реальных данных. В последнее время были предложены некоторые гибридные подходы.

Многие из существующих эвристических методов, полученные из области анализа документов, основанные на однородности цвета или интенсивности символов. Они обнаруживают области символов в изображении, а затем группируют их в слова и строки текста на основе геометрических ограничений. Эти методы, известные также как методы связных компонент [1-3], удовлетворительно работают только на высококачественных изображениях с простым фоном при известном цвете текста, что не имеет места в случае с видеоизображениями. Кроме того, текст в видеоизображениях часто искажается в цвете из-за сжатия видео.

Некоторые другие эвристические методы обнаруживают текст на основании информации о контурах. В [4] применяется горизонтально дифференцирующий фильтр, размером 3×3 пикселя, к исходному изображению с последующей бинаризацией с соответствующим порогом. В [5] предлагают метод, основанный на карте контуров, созданной с помощью оператора Собеля, которая сглаживается с помощью специальных фильтров и подвергается обработке с использованием морфологических операций. В [6] и [7] предлагается использование локального порога на карте интенсивности контура, созданной с помощью оператора Собеля. В [8] используется карта контуров, полученная с помощью детектора Канни, морфологические операции и анализ проекций обнаруженных областей на оси координат. В [9] вместо использования карты контуров для обнаружения текстовых областей предлагается использование карты перехода, сгенерированной на основании изменения интенсивности и насыщенности. Для проверки принадлежности пикселя к текстовой области применяется эвристическое правило, основанное на использовании различных локальных бинарных образов. Эти эвристические методы доказали свою эффективность для конкретных приложений с высококонтрастными изображениями и относительно однородным фоном. Однако их использование ограничивается тем, что много параметров должны быть оценены экспериментально.

Целью данной работы является разработка алгоритма и методики обнаружения текстовых областей на изображениях с неоднородным фоном.

Обнаружение текстовых областей

Видеопоследовательность рассматривается как набор кадров (отдельных изображений). Обрабатываемое изображение может иметь ряд особенностей: низкое разрешение, различный размер шрифтов текста на одном изображении, деформация надписей (поворот, расположение надписи по дуге, изменение размера символов от начала к концу надписи), слабый контраст между текстом и фоном, неоднородный фон с резкими перепадами, блики источников света на поверхностях объектов и т.д., что затрудняет обнаружение текстовых областей. Для обнаружения текстовых областей с использованием градиентных методов мы используем подход, базирующийся на работах [8], [10]. Этот подход основывается на том, что текстовые строки имеют сильные вертикальные контуры, расположенные в горизонтальном направлении с высокой плотностью. Использование информации о контуре для обнаружения текстовых областей оправдано тем, что для читабельности текст должен иметь достаточно четкие контуры.

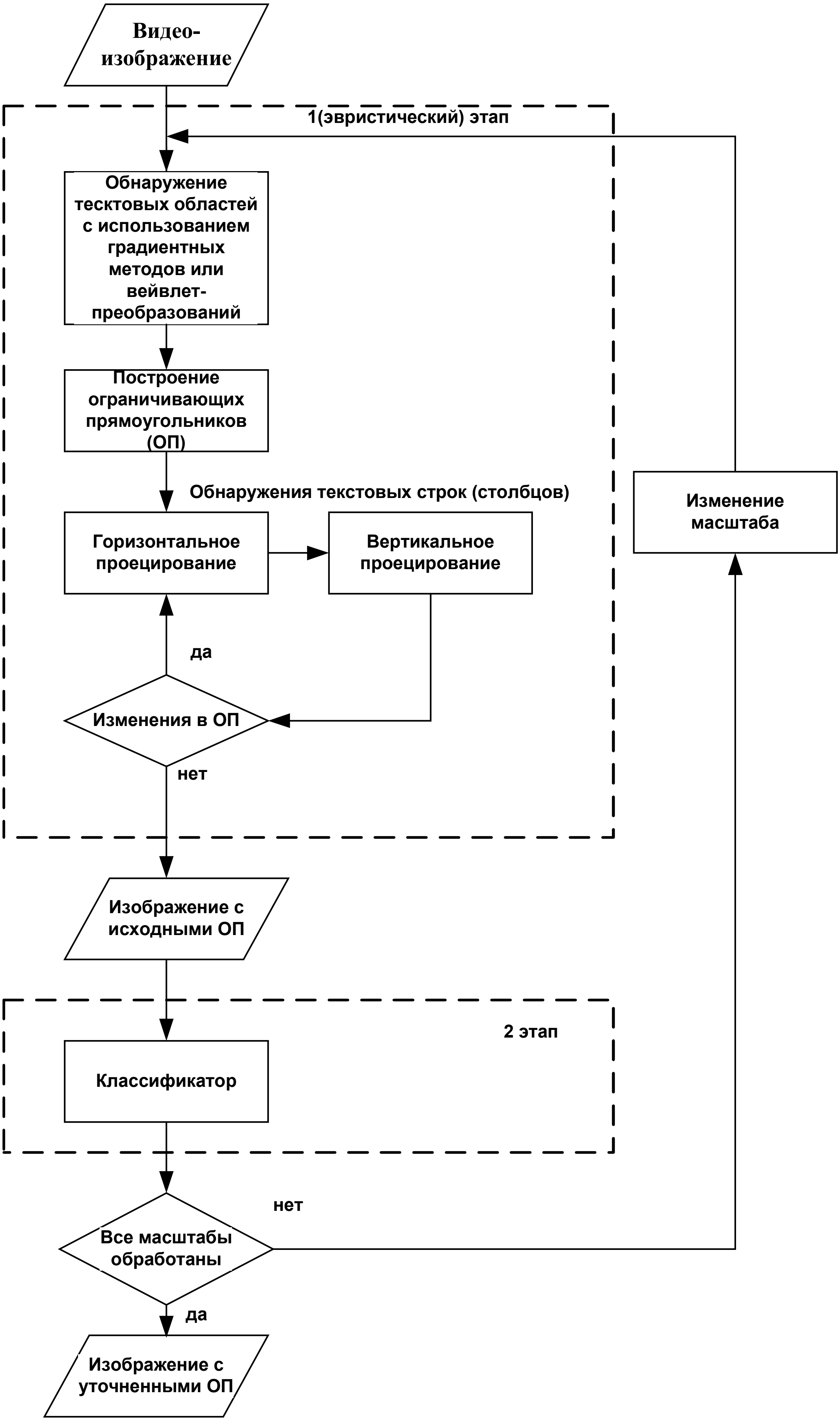
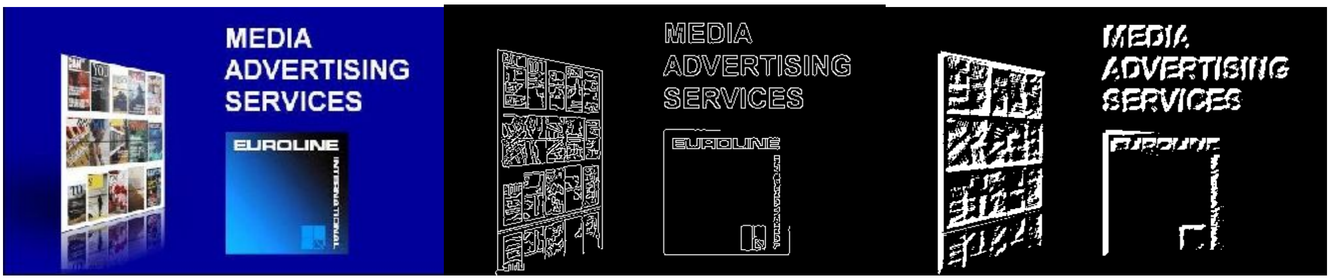


Рисунок 1 – Граф-схема алгоритма обнаружения текстовых областей

В работе предлагается двухэтапная схема и алгоритм обнаружения текстовых областей в видеопоследовательностях (рис. 1), основанные на методах и алгоритмах, предложенных в [10], и рассматривается первый этап.

На первом (эвристическом) этапе обнаружения текстовых областей в видеопоследовательностях реализованы алгоритмы обнаружения текстовых областей и текстовых строк (столбцов) с использованием градиентных методов или вейвлет-преобразований на основе анализа перепадов интенсивности в локальных областях изображения. В результате осуществляется отбор областей изображений, в которых может быть текстовая информация.

Для получения карты контуров исходного изображения (рис. 2а) преобразовываем его в полутоновое изображение и используем оператор Канни (рис. 2б). В идеале получаем изображение с набором единичных пикселей контура (белых) и остальных пикселей, равных нулю (черных). В дальнейшем используем пространственную фильтрацию контурного изображения с помощью скользящей квадратной маски (рис. 2в).



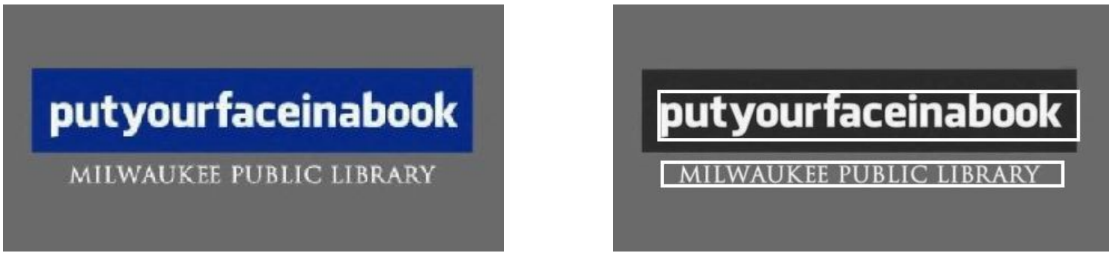
а) б) в)

Рисунок 2 – Пример обнаружения текстовых областей с использованием градиентных методов: а) исходное изображение; б) контурное изображение по методу Канни; в) контурное изображение после фильтрации с помощью скользящей маски

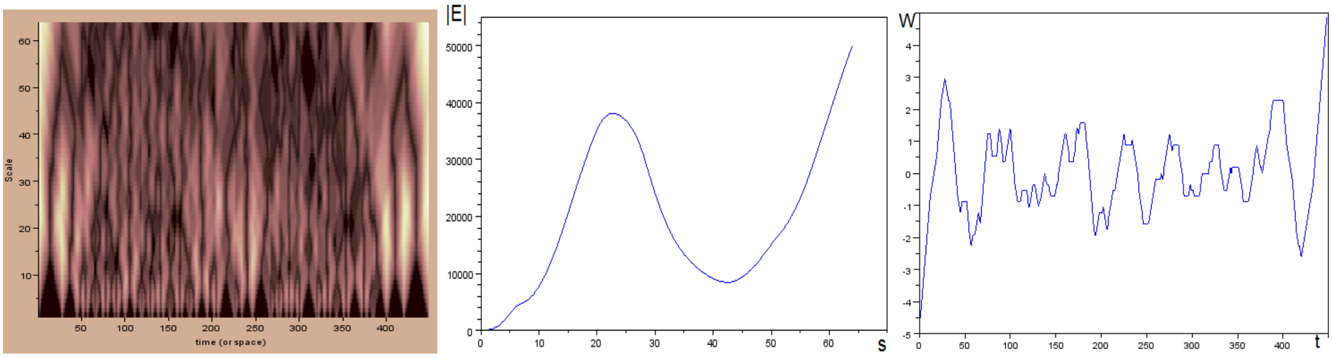
Один из перспективных подходов к определению свойств объектов на изображении основывается на результатах мультимасштабного контурного анализа, который может проводиться с разной детальностью в зависимости от требований к извлекаемой информации. Для мультимасштабного контурного анализа обычно используют непрерывное вейвлет-преобразование (ВП) [11] с действительными симметричными нечетными вейвлет-функциями (ВФ), $\psi(x) = -\psi(-x)$, которые имеют компактный или эффективный носитель, в частности, вейвлеты Гаусса, базисные функции гиперболического вейвлет-преобразования (ГВП) и др. [11-13]. Методика автоматического выбора масштаба на основе анализа интегральной функции распределения энергии по масштабам вейвлет-преобразования (ИРЭМ) предложена в [13]. Масштаб выбирается как абсцисса точки последнего минимума (или точки последнего перегиба) кривой ИРЭМ. Если минимум или перегиб отсутствуют в рассматриваемом диапазоне масштабов, то в данной строке нет текстовой информации, и она исключается из дальнейшего анализа.

Для локализации объектов на изображении, как правило, применяют построчную (либо постолбцовую) обработку, что позволяет обрабатывать функцию интенсивности в строке (столбце) как одномерный сигнал. Для расширения пространственной локализации и подчеркивания перепадов интенсивности могут применяться действительные вейвлеты Гаусса, Хаара, базисные функции гиперболического вейвлет-преобразования и т.д. [11-13].

Пример обнаружения текстовых областей с использованием вейвлет-преобразований представлен на рис. 3.



а) б)



в) г) д)

Рисунок 3 – Пример обнаружения текстовых областей с использованием вейвлет-преобразования Хаара: а) исходное изображение; б) полученный результат; в) ВП строки изображения с масштабами от 1 до 64; г) график функции ИРЭМ |E|=*f*(*s*), s – масштаб ВП; д) ВП строки изображения с масштабом 33

Для улучшения отбора текстовых областей на первом этапе после выделения контуров выполняется морфологическая операция наращивания областей, чтобы объединить контуры символов каждой текстовой строки. При этом используется горизонтальный структурный элемент, размер которого зависит от предполагаемого максимального расстояния между символами. Также используется морфологическая операция «открытие» для удаления шума в предполагаемых текстовых областях. Полученные на данном этапе области вписываются в прямоугольник (рис. 4) и являются кандидатами для проверки на принадлежность к тексту на втором этапе.



а) б)

Рисунок 4 – Пример обнаружения текстовых областей с использованием градиентных методов: а) морфорлогическое открытие изображения; б) полученный результат

В дальнейшем для устранения шума, отделения строк и столбцов текста внутри прямоугольной области используем метод горизонтальных и вертикальных проекций. Выполняем соответствующее проецирование и отбрасываем строки (столбцы), для которых полученная сумма не превышает заданного порога. Затем считаем, что вертикально разделенные части области остаются связанными, если расстояние между ними

меньше порога, который зависит от высоты *h* текстовой строки кандидата (обычно он равен 1.5*h*). Данная процедура вертикального и горизонтального проецирования повторяется до тех пор, пока не будет достигнуто устойчивое состояние.



Рисунок 5 – Примеры обнаружения текстовых строк (столбцов)

При использовании вейвлет-преобразования на первом этапе при сложном и неоднородном фоне и различных размерах шрифтов текста автоматический выбор масштаба может быть не эффективным. В этом случае производится обработка исходного изображения с несколькими (обычно от 3 до 5) различными масштабами (на рис. 1 этому соответствует ветвь с оператором «изменение масштаба») и полученные результаты объединяются. Такой подход позволяет обнаруживать текстовые области с существенно различающимися размерами шрифта, что в целом повышает качество обнаружения.

Экспериментальные результаты

Для экспериментов использовалась выборка, состоящая из набора более 200 изображений, полученных из новостных или рекламных телевизионных роликов, а также изображений вагонов, автомобилей и т.д. Пример обнаружения текста представлен на рис. 6.



Рисунок 6 – Примеры обнаружения текстовых областей

Для оценки точности обнаружения текстовых областей использованы меры точности и отзыва [10].

Мера точности (Precision) определяется

$$\text{Precision} = (\text{OTP} / \text{OP}) \times 100\%,$$

где OTP – количество обнаруженных текстовых областей; OP – обнаруженные области (текстовые, нетекстовые).

Мера отзыва Recall определяется

$$\text{Recall} = (\text{OTP} / (\text{OTP}+\text{HOTP})) \times 100\%,$$

где OTP – количество обнаруженных текстовых областей; HOTP– количество текстовых областей, которые не обнаружены.

Результаты оценки точности обнаружения текстовых областей представлены в табл. 1.

Таблица 1 – Оценки точности обнаружения текстовых областей

Название алгоритма	Мера точности (%)	Мера отзыва (%)
Градиентный	58	87
Градиентный с морфологией	59	90
С использованием ВП	78	91

Полученные результаты показывают, что использование ВП даже на первом этапе обеспечивает точность обнаружения на 18% – 20% выше, чем остальные алгоритмы. Градиентный алгоритм обнаружения текстовых областей находит слишком маленькие предполагаемые текстовые области, в то время как алгоритм с использованием ВП рассматривает их как единую область. Поэтому при использовании градиентного алгоритма часто обнаруживаются нетекстовые области.

На втором этапе двухэтапной схемы обнаружения текстовых областей в видеопоследовательностях оценивается вероятность принадлежности к тексту каждого пикселя отобранных на первом этапе областей с использованием классификатора (нейронной сети, машины опорных векторов и т.д.). По результатам оценки с использованием эмпирически установленного порога принимается окончательное решение о наличии текстовой информации в данной области и уточняются ее границы. В настоящий момент продолжаются исследования по применению классификаторов на основе нейронной сети или машины опорных векторов для повышения точности распознавания текстовых областей в видеопоследовательностях.

Выводы

Двухэтапная схема обнаружения текстовых областей в видеопоследовательностях и использование вейвлет-преобразования на этапе получения контуров символов в сочетании с применением морфологических операций позволяет повысить точность локализации текстовых областей, что в конечном итоге способствует повышению эффективности систем распознавания текстовой информации.

Использование автоматического выбора масштаба вейвлет-преобразования или обработка изображения с разными масштабами вейвлет-преобразований позволяет улучшить точность обнаружения текстовых областей на неоднородном фоне.

Предложенный алгоритм и методика выделения текстовых областей на неоднородном фоне в видеопоследовательностях может быть полезна для специалистов, занимающихся компьютерной обработкой и распознаванием изображений.

Литература

1. Воскресенский М.Е. Метод параметрической оптимизации процесса принятия решений в системах распознавания текстовых меток на видеоизображениях /М.Е. Воскресенский // Компьютерная оптика. – 2009. – Т. 33, № 2. – С. 202-209.
2. Lienhart R. Automatic text segmentation and text recognition for video indexing / R.Lienhart, W. Effelsberg / Springer Multimedia Systems. – 2000. – № 8. – P. 69-81.
3. Sobottka K. Identification of text on colored book and journal covers / Sobottka K., Bunke H., Kronenberg H. // International Conference on Document Analysis and Recognition, 1999. – P. 57-63.
4. Video OCR for digital news archives / [T. Sato, T. Kanade, E. Hughes, M. Smith] // IEEE Workshop on Content-Based Access of Image and Video Databases, 1998. – P. 52-60.
5. A video text detection and recognition system, multimedia and expo / [J. Xi, X.-S. Hua, X.-R. Chen, L. Wenyin, H. Zhang] // IEEE International Conference, 2001. – P. 873-876.

6. Cai M. A new approach for video text detection / M. Cai, J. Song, M.R. Lyu // IEEE International Conference on Image Processing, 2002. – P. 117-120.
7. Lyu M.R. A comprehensive method for multilingual video text detection, localization, and extraction / M.R. Lyu, J. Song, M. Cai // IEEE Transactions. Circuit and Systems for Video Technology. – 2005. – №15 (2). – P. 243-255.
8. Anthimopoulos M. Multiresolution text detection in video frames / M. Anthimopoulos, B. Gatos, I. Pratikakis // International Conference on Computer Vision Theory and Applications, 2007. – P. 161-166.
9. Kim W. A new approach for overlay text detection and extraction from complex video scene / W. Kim, C. Kim // Image Processing, IEEE Transactions. – 2009. – № 18 (2). – P. 401- 411.
10. Anthimopoulos Marios. A two-stage scheme for text detection in video images / M. Anthimopoulos, B. Gatos, I. Pratikakis // Image and Vision Computing. – 2010. – № 28. – P. 1413-1426.
11. Малла С. Вейвлеты в обработке сигналов / Малла С. – М. : Мир, 2005. – 671 с.
12. Антощук С.Г. Анализ базисных функций вейвлет-преобразования при мультимасштабном контурном представлении изображений / С.Г. Антощук, А.А. Николенко, Е.В. Ткаченко // Електромашинобудування та електрообладнання. – 2009. – Вип. 72. – С. 15-19.
13. Использование вейвлет-преобразования при локализации последовательностей символов / [С.Г. Антощук, А.А. Николенко, Е.В. Ткаченко, О.Ю. Бабилунга] // Искусственный интеллект. – 2009. – № 4. – С. 23-29.

Literatura

1. Voskresenskij M.E. Komp'juternaja optika. 2009. Tom33.№ 2. S. 202-209.
2. Lienhart R. Springer Multimedia Systems. 2000. № 8. P. 69-81.
3. Sobottka K. International Conference on Document Analysis and Recognition. 1999. P. 57-63.
4. SatoT. IEEE Workshop on Content-Based Access of Image and Video Databases. 1998. P. 52-60.
5. Xi J. IEEE International Conference. 2001. P. 873-876.
6. Cai M. IEEE International Conference on Image Processing. 2002. P. 117-120.
7. Lyu M.R. IEEE Transactions. Circuit and Systems for Video Technology. 2005.№ 15 (2). P. 243-255.
8. Anthimopoulos M. International Conference on Computer Vision Theory and Applications. 2007. P. 161-166.
9. Kim W. Image Processing, IEEE Transactions. 2009. №18 (2). P. 401-411.
10. Anthimopoulos Marios. Image and Vision Computing . 2010. № 28. P. 1413-1426.
11. Malla S. Vejvlety v obrabotke signalov. M.: Mir, 2005. 671 s. .
12. Antoshhuk S.G. Elektromashhnobuduvannja ta elektroobladnannja. 2009. Vyp. 72. S. 15-19.
13. Antoshhuk S.G. Iskusstvennyj intellekt. 2009. № 4. S. 23-29.

RESUME

A.A.Nikolenko, Tien T.K. Nguyen

Text Region Detection in Video Frames

In this paper we consider the problem of text regions detection on a non-uniform background in video frames. It is proposed the two-stage scheme, the algorithm and the technique for detection of text regions. The first stage is described in detail.

Compared with existing algorithms and approaches, it is proposed to use to highlight the edges of characters of text continuous wavelet transform with automatic selection of scale or iterative processing at different scales at the first stage. Next post-processing with threshold and search for connected components can select a rectangular area containing the text information. Morphological operations are used to improve the quality of text region detection.

The experimental testing of base, which contains over 100 images, is made. The results show that using the wavelet transform, even at the first stage, provides detection accuracy to 18% – 20% higher than gradient algorithms.

At the second stage, classifier is used to decide if regions derived an the first stage are text regions. Currently, studies on the use of classifiers based on neural networks or support vector machines for improving the recognition accuracy of text regions in video frames are going on.

Статья поступила в редакцию 07.06.2012.